JCCES-II®

JOUVE-CEREBRALS CRYSTALLIZED EDUCATIONAL SCALE II

Technical Manual — Pilot Calibration

Cognitive Acumen Index (CAI) · pilot calibration

METHODS & EVIDENCE

2026 EDITION

CONTENTS

1 The Crystallized-Ability Construct and the Design of the Battery

2 The Pilot Calibration Study

3 Score Scale, the Six Subtests, and the CAI

4 Item Format, Answer Keys, and Calibration

5 Administration and the Testing Session

6 Scoring and Provisional Equating to the CAI

7 Reliability and Precision of Measurement
8 Validity Evidence: Internal Structure
9 Validity Evidence: Convergent Correlations
10 Interpretation of Scores
11 Development Status, Uses, and Boundaries
— References

CHAPTER 1

The Crystallized-Ability Construct and the Design of the Battery

Development edition. The JCCES-II is a research instrument under development. The evidence in this manual comes from a single pilot calibration and establishes a calibration framework rather than population norms; scores require further calibration before they can be interpreted for individual decisions (Jouve, 2024). Every chapter should be read in that light.

Crystallized ability — the store of acquired verbal and quantitative knowledge and the reasoning that operates over it — is one of the broad abilities of the Cattell–Horn–Carroll model and carries much of the variance that general-knowledge and reasoning batteries measure (Carroll, 1993; McGrew, 2009). The Jouve Cerebrals Crystallized Educational Scale II (JCCES-II) renews the first-edition JCCES as a broader, six-subtest battery of that ability, spanning linguistic, quantitative, and knowledge content in both reasoning and information formats.

The battery comprises 260 dichotomously scored items across six subtests: Comparative Linguistic Analogies (CLA, 40 items), Progressive Linguistic Sequencing (PLS, 43), Algebraic Reasoning Analysis (ARA, 44), Geometric Reasoning Analysis (GRA, 37), Verbal Humanities Information (VHI, 46), and Quantitative Scientific Information (QHI, 47); three Verbal Analogies items are retained from the first-edition

bank as anchors to its scale (Jouve, 2024). The instrument reports a single total-score index, the Cognitive Acumen Index (CAI), on the deviation-IQ metric (M = 100, SD = 15) of the first edition. Whether the six subtests can be scored as separate dimensions is an empirical question the pilot addresses (chapter 8); at the present sample size only the total score is reported. The chapters that follow adopt the documentation structure recommended by the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014), with the pilot status noted at each interpretive step.

CHAPTER 2

The Pilot Calibration Study

The evidence in this manual is drawn from one published study, a pilot calibration and equating of the JCCES-II (Jouve, 2024). That study administered the full 260-item battery to 62 examinees, established the item and reliability properties reported in chapters 4 and 7, examined the internal structure by parallel analysis and a comparison of measurement models (chapter 8), correlated the total score with external instruments (chapter 9), and equated the total to the first-edition CAI metric (chapter 6). The examinees were a self-selected, above-average testing sample; subsamples had also taken external instruments — the first-edition JCCES, the JCTI, the IAW, and WAIS-IV verbal subtests — which supply the convergent evidence. The study is explicit that it establishes a calibration framework, not population norms, and that the instrument requires further calibration before individual scores can be interpreted (Jouve, 2024). This manual reports what that pilot establishes and marks what it does not.

CHAPTER 3

Score Scale, the Six Subtests, and the CAI

The JCCES-II reports a single total-score index, the Cognitive Acumen Index, on the deviation-IQ metric (M = 100, SD = 15) inherited from the first edition through the equating of chapter 6. The six subtests

contribute their item scores to the total; they are not at present reported as separate scaled scores, because the pilot's internal-structure evidence does not yet identify them as separable dimensions at this sample size (chapter 8). Table 3.1 lists the subtests and their content.

Table 3.1

The six JCCES-II subtests

Subtest	Items	Content and format
Comparative Linguistic Analogies (CLA)	40	Verbal analogical reasoning
Progressive Linguistic Sequencing (PLS)	43	Verbal sequential reasoning
Algebraic Reasoning Analysis (ARA)	44	Quantitative — algebraic reasoning
Geometric Reasoning Analysis (GRA)	37	Quantitative — geometric reasoning
Verbal Humanities Information (VHI)	46	Acquired verbal knowledge
Quantitative Scientific Information (QHI)	47	Acquired quantitative knowledge

Note. The six subtests total 260 dichotomous items; three Verbal Analogies anchor items retained from the first edition are not part of the six-subtest count (Jouve, 2024). The total raw score S is the sum of correct responses.

Reporting the total on the first-edition CAI metric is a deliberate choice: it fixes the unit and origin of the scale by anchoring the new battery to an existing, interpreted index rather than to the pilot sample's own mean and spread. Because the anchoring is a provisional pilot equating (chapter 6), a JCCES-II CAI should be read as an estimate on that metric pending fuller calibration, not as a normed standard score.

CHAPTER 4

Item Format, Answer Keys, and Calibration

Each JCCES-II item is scored dichotomously, 1 for a correct response and 0 otherwise, and the 260 item scores sum to the total raw score. To protect item security, no item text or key entry is reproduced in this manual.

IRT calibration

The items were calibrated in the pilot under a two-parameter logistic model, estimating a difficulty and a discrimination for each item (Jouve, 2024; Embretson & Reise, 2000). Under the 2PL the probability of a correct response rises with ability according to

$$P(x_i = 1 \mid \theta) = \frac{1}{1 + \exp[-a_i\,(\theta - b_i)]},$$

with θ the examinee's ability, b_i the item difficulty, and a_i the item discrimination. Every subtest returned positive item discriminations, and the total raw scores ranged from 51 to 236 (M = 144.8, SD = 41.26), a wide spread that supports estimation across the ability range sampled by the pilot (Jouve, 2024). Because the calibration rests on 62 examinees, the item parameters are provisional and are expected to be re-estimated as the calibration sample grows.

CHAPTER 5

Administration and the Testing Session

The JCCES-II is administered online as a self-paced session in which the examinee works through the 260 items across the six subtests. The battery is designed for literate adults; the reasoning subtests require no specialized training beyond ordinary literacy and elementary quantitative fluency, and the information subtests draw on acquired verbal and scientific knowledge.

Administration is untimed in the sense that the total is defined on completed responses rather than on a rate, so that the CAI reflects acquired knowledge and reasoning accessible without time pressure. As with the other Cogn-IQ instruments, a valid score assumes an attentive, uninterrupted session completed without external reference aids, since the information subtests in particular could be looked up. Because the

instrument is in pilot calibration, its scores are intended for research and development use rather than for individual high-stakes decisions.

CHAPTER 6

Scoring and Provisional Equating to the CAI

Scoring proceeds in two steps. The total raw score S is formed as the sum of the 260 dichotomous item scores. That total is then converted to the Cognitive Acumen Index by equating it onto the first-edition CAI metric.

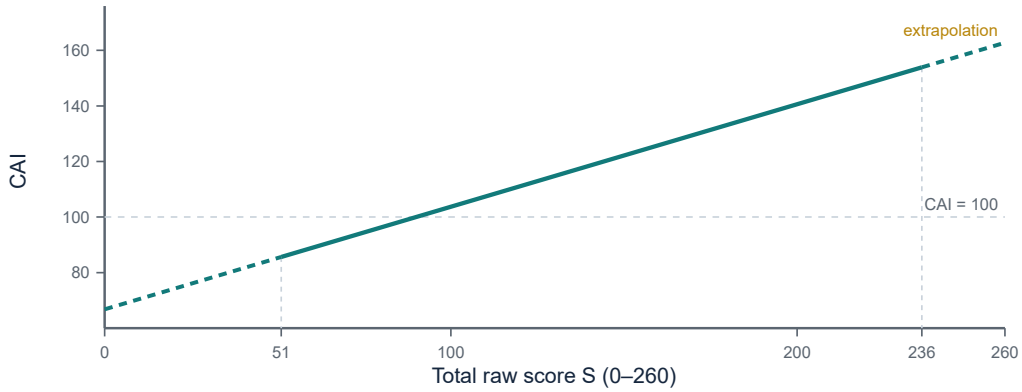
The pilot equating is a linear mean-and-sigma map from the total raw score to the CAI, obtained by regressing the first-edition CAI on the JCCES-II total in the subsample that took both,

$$\mathrm{CAI} = 66.8 + 0.369\,S,$$

which fitted with $r = .906$ and is offered by the source study as a provisional calibration anchor rather than a final norm (Jouve, 2024; Kolen & Brennan, 2014). Operationally the platform delivers the CAI through a fixed monotone polynomial in the raw score that reproduces this linear anchor across the range of totals observed in the pilot (51 to 236), diverging from the straight line only in extrapolation above the highest observed score, where the pilot supplies no data (Figure 6.1). A CAI computed for a total beyond the pilot's observed maximum is therefore an extrapolation and should not be interpreted as a calibrated score. As the calibration sample grows, both the linear anchor and its operational conversion are expected to be revised.

Figure 6.1

Provisional equating of the JCCES-II total raw score to the CAI metric



Note. Solid segment spans the pilot's observed raw-score range (51–236), over which the operational conversion reproduces the linear anchor CAI = 66.8 + 0.369S (r = .906; Jouve, 2024). Dashed segments are extrapolation beyond the observed data.

CHAPTER 7

Reliability and Precision of Measurement

Internal consistency was strong throughout the pilot. On the 260-item total the KR-20 coefficient reached .978, and each of the six subtests was internally consistent in its own right, with KR-20 coefficients ranging from .881 to .949 (Jouve, 2024; Kuder & Richardson, 1937). The item statistics support this: median item-rest correlations ran from .37 to .61 across the subtests, and no subtest showed the weak item-total relations that would signal construct-irrelevant content (Jouve, 2024; Nunnally & Bernstein, 1994). Table 7.1 summarizes the coefficients.

Table 7.1

Internal consistency in the pilot calibration (N = 62)

Score	Items	KR-20
CAI total	260	.978
Six subtests (range)	37–47	.881–.949

Note. KR-20 from the pilot calibration (Jouve, 2024). Median item-rest correlations ranged from .37 to .61 across subtests. The standard error of a score on the IQ metric is $\mathrm{SEM} = \mathrm{SD}\sqrt{1-\rho}$*; at a total-score reliability near .98 this is*

about two IQ points, an internal-consistency figure to be read alongside the provisional status of the equating.

These are internal-consistency coefficients estimated on the calibration sample; they describe how coherently the items measure a common dimension, not the temporal stability of the score, which the pilot did not assess. Because the coefficients rest on 62 examinees, they are reported as pilot values to be confirmed as the sample grows.

CHAPTER 8

Validity Evidence: Internal Structure

Whether the six subtests measure one crystallized dimension or several is the central structural question, and the pilot examined it by parallel analysis, exploratory factoring, and a maximum-likelihood comparison of four measurement models — a one-factor model, the a priori hierarchical (three-factor) model, and content-domain and process alternatives — with the models compared by fit-aware AIC and BIC and likelihood-ratio tests and the exact-fit statistic recalibrated by a Bollen–Stine bootstrap (Jouve, 2024; Horn, 1965; Bollen & Stine, 1992).

The evidence points in a consistent direction while stopping short of licensing subtest scores. Parallel analysis retained a single factor. In the confirmatory comparison the a priori three-factor structure fit best in absolute terms, but its group factors are defined by only two indicators each and are only weakly identified at this sample size, and the amount of reliable variance they carry beyond the general factor is modest (general-factor ECV = .69, omega-hierarchical ω_h = .82; interfactor correlation .76; Jouve, 2024; Rodriguez et al., 2016). Table 8.1 collects the structure indices. Taken together — a single factor retained by parallel analysis, a dominant general factor in the bifactor decomposition, and group factors too weakly identified to score — the evidence supports reporting only the total CAI at this stage, with the multi-factor structure a hypothesis to be tested as the calibration sample grows.

Table 8.1

Internal-structure indices from the pilot calibration

Index	Value	Reading
Parallel analysis	1 factor	One dimension retained
General-factor ECV	.69	Share of common variance on the general factor
Omega-hierarchical (ω_h)	.82	Reliable variance due to the general factor
Interfactor correlation	.76	Group factors highly correlated

Note. From the pilot measurement-model comparison (Jouve, 2024). The a priori three-factor model fit best in absolute terms, but its two-indicator group factors are weakly identified at N = 62, so only the total score is reported.

CHAPTER 9

Validity Evidence: Convergent Correlations

The pilot correlated the JCCES-II total with a profile of external cognitive measures taken by overlapping subsamples (Jouve, 2024). The strongest relations were with the first-edition JCCES, as instrument renewal predicts: the total correlated .91 with the first-edition Cognitive Acumen Index and .89 with its Mathematical Problems subtest, and the correlation with the CAI is what the equating of chapter 6 rests on. Relations with a broader inductive-reasoning index (JCTI, r = .82) and a vocabulary index (IAW-VPI, r = .80) were also substantial, consistent with the overlap of crystallized ability with reasoning and lexical knowledge near the top of the cognitive hierarchy (Carroll, 1993). Table 9.1 gives the observed coefficients with their 95% confidence intervals and their range-restriction-corrected counterparts.

Table 9.1

Convergent correlations of the JCCES-II total with external measures

External measure	*n*	*r* [95% CI]	*r* corrected [95% CI]
JCCES-I Cognitive Acumen Index (CAI)	25	.91 [.80, .96]	.88 [.75, .94]
JCCES-I Verbal Acumen Index (VAI)	20	.80 [.55, .92]	.78 [.53, .91]
JCCES-I Mathematical Problems (scaled)	25	.89 [.77, .95]	.87 [.73, .94]
JCCES-I Verbal Analogies (scaled)	25	.78 [.56, .90]	.77 [.55, .89]
JCCES-I General Knowledge (scaled)	25	.73 [.47, .87]	.80 [.56, .91]
JCTI Inductive Reasoning Index (IRI)	14	.82 [.51, .94]	.77 [.45, .92]
IAW Vocabulary Proficiency Index (VPI)	27	.80 [.60, .90]	.73 [.52, .86]

Note. Observed Pearson correlations of the JCCES-II total with each criterion, with 95% confidence intervals; corrected coefficients apply Thorndike's Case II range-restriction correction, which can move a correlation in either direction depending on the selection variable (Thorndike, 1949; Sackett & Yang, 2000; Jouve, 2024). The interval widths carry the precision of each pilot subsample.

The confidence intervals are the point to hold onto: the convergent pattern is coherent and points to a crystallized-ability interpretation, but each coefficient is estimated with the precision its pilot subsample allows, and the intervals are correspondingly wide. The correlations are also observed on a self-selected, above-average sample and are attenuated by imperfect reliability in both measures (Spearman, 1904); the corrected column addresses restriction of range but not attenuation. The validity evidence is therefore best read as a promising but preliminary calibration, to be firmed up as the samples grow.

CHAPTER 10

Interpretation of Scores

The JCCES-II reports a single CAI on the deviation-IQ metric of the first edition, and its interpretation is governed by the instrument's development status. Because the equating is a provisional pilot anchor rather than a norm, a JCCES-II CAI is an estimate on the first-edition metric and should be interpreted only as such, with the measurement error of chapter 7 and the calibration uncertainty of chapter 6 both in view. Table 10.1 gives the conventional descriptive bands for the deviation-IQ metric to which the CAI is anchored; they carry the meaning they carry on the first edition, subject to the calibration caveat.

Table 10.1

Descriptive classification of CAI scores

CAI range	Classification	Approx. percentile
130 and above	Very high	98 and above
120–129	High	91–97
110–119	High average	75–90
90–109	Average	25–74
80–89	Low average	9–24
Below 80	Below average	Below 9

Note. Percentiles are those of the standard normal distribution at the stated CAI bounds (M = 100, SD = 15) and are approximate. The bands describe the metric to which the pilot equating anchors the CAI; they do not certify JCCES-II population norms, which the pilot does not establish.

Two cautions are decisive at this stage. First, the CAI is reported only as a total; the subtest profile is not scored, because the pilot does not yet identify the subtests as separable dimensions (chapter 8). Second, a CAI is a provisional, research-use estimate: it should not be used on its own for individual high-stakes decisions until the calibration is confirmed on larger samples.

Development Status, Uses, and Boundaries

The JCCES-II is a research instrument in pilot calibration. Its evidence — strong internal consistency, a wide item-difficulty range, a dominant general factor, and substantial convergent relations with established crystallized and reasoning measures — is consistent with a broad crystallized-ability battery and supports its continued development. What that evidence does not yet establish should be stated as plainly.

Three boundaries define the appropriate use of the instrument today. First, the norming is provisional: the CAI rests on a pilot equating to the first-edition metric, not on a representative standardization sample, so the instrument does not yet carry population norms (Jouve, 2024; Kolen & Brennan, 2014). Second, the evidence rests on a single pilot of 62 examinees and overlapping subsamples for the criterion coefficients, so the item parameters, reliability, structure, and validity coefficients are all pilot estimates to be re-established as the calibration sample grows — a limitation the confidence intervals of chapter 9 make concrete. Third, only the total score is interpretable; the six-subtest structure is a design hypothesis not yet confirmed at scale. Within these boundaries the JCCES-II is appropriately used for research and continued calibration, and a reported CAI should be read as a provisional estimate of crystallized ability on the first-edition metric, not as a normed general intelligence quotient.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Bollen, K. A., & Stine, R. A. (1992). Bootstrapping goodness-of-fit measures in structural equation models. *Sociological Methods & Research, 21*(2), 205–229. https://doi.org/10.1177/0049124192021002004

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press. https://doi.org/10.1017/CBO9780511571312

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Lawrence Erlbaum Associates.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*(2), 179–185. https://doi.org/10.1007/BF02289447

Jouve, X. (2024). *Jouve Cerebrals Crystallized Educational Scale II: Pilot calibration and equating to the JCCES-I Cognitive Acumen Index*. Cogn-IQ Research Papers. https://pubscience.org/ps-1mVDO-9b28f7-uQHF

Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices* (3rd ed.). Springer. https://doi.org/10.1007/978-1-4939-0317-7

Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika, 2*(3), 151–160. https://doi.org/10.1007/BF02288391

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project. *Intelligence, 37*(1), 1–10. https://doi.org/10.1016/j.intell.2008.08.004

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods, 21*(2), 137–150. https://doi.org/10.1037/met0000045

Sackett, P. R., & Yang, H. (2000). Correction for range restriction: An expanded typology. *Journal of Applied Psychology, 85*(1), 112–118. https://doi.org/10.1037/0021-9010.85.1.112

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology, 15*(1), 72–101. https://doi.org/10.2307/1412159

Thorndike, R. L. (1949). *Personnel selection: Test and measurement techniques*. Wiley.